



U.S. Department of Energy

Oakland Operations Office, Oakland, California

RECORD COPY
PROJECT: 17823
CAT. 13.9-2.6

FINAL DATA INTEGRATION STRATEGY MEMORANDUM

for the:

LABORATORY FOR ENERGY-RELATED HEALTH RESEARCH
UNIVERSITY OF CALIFORNIA, DAVIS

Prepared for:

United States Department of Energy
Oakland Operations Office
1301 Clay Street
Oakland, California 94612-5208

Prepared by:

Weiss Associates
5801 Christie Avenue, Suite 600
Emeryville, CA 94608-1827

May 31, 2000
Rev. 0

DOE Oakland Operations Contract DE-AC03-96SF20686



ENVIRONMENTAL HEALTH & SAFETY
UNIVERSITY OF CALIFORNIA
ONE SHIELDS AVENUE
DAVIS, CALIFORNIA 95616-8648
(530) 752-1493 FAX (530) 752-4527

June 7, 2000

Ms. Kathy Setian
U.S. Environmental Protection Agency
Federal Facilities Cleanup Office, Section I
75 Hawthorne Street
Mail Code SFD-72
San Francisco, CA 94105

Re: Final Data Integration Strategy Memorandum

Enclosed please find one copy of the Final Data Integration Strategy Memorandum (DISM) prepared by the U.S. Department of Energy (DOE) and the University of California, Davis (UC Davis). This memorandum has been prepared to provide a framework by which existing project databases can be efficiently integrated for future uses by DOE and UC Davis. Currently, the anticipated future uses include limited data exchanges between DOE and UC Davis to support site remedies and integration of all site-wide data to support the risk assessment scheduled in 2003.

The enclosed plan has been reviewed by DOE, UC Davis and their respective contractors and deemed comprehensive to address the needs of the site data. By signature below, DOE and UC Davis agree to implement the planned data integration process described in the DISM.

If you have any questions about the enclosed document, please call me at (530) 752-6041, or write me at the above address.

Handwritten signature of Brian Oatman in black ink.

Brian Oatman
University of California, Davis

Handwritten signature of Rich Fallejo in black ink.

Rich Fallejo
U.S. Department of Energy

Ms. Kathy Setian
June 7, 2000

Page 2

Attachments: Final Data Integration Strategy Memorandum

cc: Susan Timm, CVRWQCB
Duncan Austin, DTSC
Sudana Kwok, DOHS
Bob Devany, Weiss Associates
Joseph Turner, URS/Dames & Moore
Julie Roth, DSCSOC
Mary Rust, DSCSOC
G. Fred Lee, G. Fred Lee & Associates

FINAL DATA INTEGRATION STRATEGY
MEMORANDUM

for the:

LABORATORY FOR ENERGY-RELATED HEALTH RESEARCH
UNIVERSITY OF CALIFORNIA, DAVIS

Prepared for:

United States Department of Energy
Oakland Operations Office
1301 Clay Street
Oakland, California 94612-5208

Prepared by:

Weiss Associates
5801 Christie Avenue, Suite 600
Emeryville, CA 94608-1827

May 31, 2000
Rev. 0

DOE Oakland Operations Contract DE-AC03-96SF20686

Issued To: _____ Date: _____

Copy No.: _____ Controlled Uncontrolled

CONTENTS

| | | |
|----|--|-----|
| 1. | INTRODUCTION | 1-1 |
| | 1.1 Objective | 1-1 |
| | 1.2 Background | 1-1 |
| | 1.3 Integration Process Overview | 1-2 |
| 2. | DATABASE FIELDS DEFINITION | 2-1 |
| 3. | DATA DICTIONARY DEVELOPMENT | 3-1 |
| 4. | DATABASE CROSS-REFERENCE STRUCTURE DEVELOPMENT | 4-1 |
| 5. | DATA SHARING | 5-1 |
| | 5.1 Points of Contact | 5-1 |
| | 5.2 Data Requests | 5-1 |
| | 5.3 Routine Data Exchange | 5-1 |
| | 5.4 Data Structure Changes | 5-2 |
| | 5.5 Exchange Medium | 5-2 |
| 6. | IMPLEMENTATION SCHEDULE | 6-1 |
| 7. | REFERENCES | 7-1 |

FIGURE

Figure 1. Data Integration Process Flowchart

APPENDIX

Appendix A. Proposed Outline for Data Dictionary

ACRONYMNS

| | |
|----------|---|
| URS/D&M | URS/Dames & Moore |
| DISM | Data Integration Strategy Memorandum |
| DOE | U.S Department of Energy |
| EDD | Electronic Data Deliverable |
| EPA | U.S. Environmental Protection Agency |
| ER | Environmental Restoration |
| LEHR | Laboratory for Energy-related Health Research |
| MOA | Memorandum of Agreement |
| PNNL | Pacific Northwest National Laboratory |
| QA/QC | Quality Assurance/Quality Control |
| RPMs | Remedial Project Managers |
| SCDS | South Campus Disposal Site |
| UC Davis | University of California at Davis |
| WA | Weiss Associates |

1. INTRODUCTION

1.1 Objective

The primary objective of this Data Integration Strategy Memorandum (DISM) is to provide a framework by which the existing project databases can be easily accessed and efficiently integrated by the U.S Department of Energy (DOE) and the University of California at Davis (UC Davis) for meeting the data needs of the LEHR/SCDS CERCLA project, including the Site-wide Risk Assessment scheduled for 2003.

1.2 Background

Under the Memorandum of Agreement (MOA) between UC Davis and DOE, primary responsibility for remediation of the LEHR/SCDS site (the Site) has been divided between UC Davis and DOE, with UC Davis accepting the lead in remediating its former disposal areas and ground water and DOE accepting the lead in remediating its former disposal area, waste treatment systems, and dog pen areas. As a result of these divided responsibilities, two separate environmental restoration (ER) databases currently exist and are maintained for the Site. The respective ER databases contain primarily data for air, soil, surface, and ground water.

The UC Davis database is currently maintained by its ER contractor, URS/Dames & Moore (URS/D&M), at their Sacramento, California, office, and the DOE database is currently maintained by its ER contractor, Weiss Associates (WA), at their Emeryville, California, office. The software used for the UC Davis database is FoxPro and for the DOE database is Access 97. The databases are stored on the respective contractors' internal computer networks. Only the URS/D&M and WA database managers can add to or change data in their respective database; however, other URS/D&M and WA staff also have access to the data in the respective databases. At present, only URS/D&M staff have access to the UC Davis database and only WA staff have access to the DOE database. However, data have been exchanged between databases upon request from DOE and/or UC Davis.

Many features are common to the two databases since they originated as one database that was maintained by URS/Dames & Moore (URS/D&M) from late 1990 through 1994. URS/D&M has continued to maintain this database for UC Davis since 1994. However, the DOE database was maintained by Pacific Northwest National Laboratory (PNNL) between late 1994 and 1996, and has been maintained by WA since then. Changes have been made to both the UC Davis and DOE databases since 1994, so the two databases no longer have identical structure. In addition to these differences, one or both databases may not presently include all the data fields potentially needed for future Risk Assessment and other project tasks.

Since the MOA was established, the UC Davis and DOE ER activities have been essentially mutually exclusive. Therefore, data collected by UC Davis/URS/D&M and maintained in the UC Davis database is used almost exclusively by UC Davis/URS/D&M, and similarly, data collected by DOE/WA and maintained in the DOE database is used almost exclusively by DOE/WA. Although there may be need to exchange limited data sets in the future, the only need for a combined database that includes all UC Davis and DOE Site data is the Site-wide Risk Assessment scheduled for 2003. Upon special request, a few data sets have been exchanged between URS/D&M and WA during the past several years. These activities have revealed that the two databases are quite similar and that limited data exchanges can be implemented with little difficulty.

1.3 Integration Process Overview

As described above, the purpose of this DISM is to ensure that data from the two databases can be exchanged and integrated as needed to meet project data needs now and throughout the foreseeable life of the ER projects. Currently, the foreseeable data exchange/integration needs are: 1) limited data exchanges between now and 2003; and, 2) integration of all data for the Site-wide Risk Assessment in 2003. The approach described in this DISM is for UC Davis and DOE to continue maintaining their independent LEHR databases while establishing a data exchange and integration process that will allow for limited as-needed data exchanges during the next three years and will also ensure smooth integration of all necessary data for the Site-wide Risk Assessment in 2003. This is the best approach because: 1) the Site areas and/or media represented in the UC Davis database are essentially independent of those represented in the DOE database, and will continue to be for the next three years; 2) maintaining independent UC Davis and DOE databases ensures the necessary quality control and responsiveness for project data needs; and, 3) data exchanges conducted to date have gone smoothly.

Figure 1 outlines the data integration process to achieve the objective stated in Section 1.1. The major steps in this process are:

1. Define database field needs;
2. Document data structures in Data Dictionaries and exchange dictionaries;
3. Refine Data Dictionaries and develop Cross-reference Structures using sample data set; and,
4. Exchange data as needed.

Each of the basic steps is presented in the following sections.

2. DATABASE FIELDS DEFINITION

Upon approval of the data integration process, URS/D&M and WA Database Managers and other appropriate technical staff will meet to discuss database fields definition. The meeting(s) will focus on data needs related to the Site-wide Risk Assessment and Data Quality Objectives. The purpose of the meeting(s) will be to ensure that both databases contain the information necessary for the Site-wide Risk Assessment, and that the information is in a format that allows for smooth integration of the two databases. Examples of database issues that might be discussed during the meeting(s) are:

- how data that represents something eliminated through removal actions is denoted;
- how dilution, duplicate analysis, duplicate sample, and split sample results are denoted;
- how level of data quality/data validation is denoted;
- how spatial locations of samples are denoted; and,
- what other physical data are stored in database and how.

The Database Managers and other technical staff will also discuss Data Dictionary requirements during this meeting. The Database Managers will determine the elements of the Data Dictionary that are necessary for efficient cross-referencing and use of the UC Davis and DOE databases by both parties. Appendix A is a proposed Data Dictionary outline for use as a starting point in discussing the necessary elements. As part of these discussions, the Database Managers will also try to identify database changes that can be easily implemented by DOE and/or UC Davis, with the goal of limiting the Cross-reference Structures (see Section 4) needed to integrate the databases.

WA will take the lead in scheduling the data needs discussions. More than one meeting may be necessary to agree on all issues. Once agreement is reached, the WA Database Manager will prepare a memo summarizing the decisions made during these meetings and present a Draft Data Dictionary outline. UC Davis will review and provide comments, which will be addressed in a follow-up memo and Final Data Dictionary outline prepared by WA.

3. DATA DICTIONARY DEVELOPMENT

Database integration will require that both parties develop, finalize, and implement Data Dictionaries that document data structure. As described above, WA and URS/D&M will discuss and verbally agree on the elements needed in the Data Dictionaries. Based on these discussions, WA will develop a Draft Data Dictionary outline for review and comment by URS/D&M, and incorporate URS/D&M comments in a Final Data Dictionary outline. This outline will be used independently by URS/D&M and WA to develop Data Dictionaries for their respective databases. At a minimum, the Data Dictionaries will include the elements described below. Both URS/D&M and WA have existing Data Dictionaries that include at least some of these elements; and these existing dictionaries will be used as a starting point to the extent possible.

Relational Information – The first part of the Data Dictionary should detail the relational structures used to store information, including:

- the database software being used;
- the information the database provides, i.e. describe the database function;
- which tables are used to store the pertinent information;
- the relationships between these tables;
- which columns or fields are used to store the information in these tables; and,
- the constraints and characteristics of the fields that are used to control the format of the stored information.

Data Characteristics – This part of the Data Dictionary should explain the interpretive information necessary for understanding the data. This section should explain naming conventions and define acronyms. It may be necessary to have a separate subsection for each type of environmental media (e.g., soil, water, and air). Data characteristics should also provide a key for:

- CAS numbers;
- CAS names;
- data validation flags;
- analytical methods;
- operable unit areas;
- field locations;
- units of measurement; and,
- sample collection type, if appropriate (i.e., grab composite, boring).

A key element of the Data Dictionary is the database field descriptions. The data dictionaries will present the current field descriptions for both the UC Davis and DOE databases. These field descriptions are already quite similar, suggesting that the Data Dictionary completion and Cross-reference Structure development should go smoothly and require a limited effort.

4. DATABASE CROSS-REFERENCE STRUCTURE DEVELOPMENT

As described in Section 2, efforts will be made to align the two databases to the extent practical prior to developing the Data Dictionaries. However, it will not be practical to align all the historical data from the two databases into identical formats. In addition, there may be database differences that are not practical to align for future data. Therefore, a Cross-reference Structure must be developed that describes the differences and correlates information between the two databases.

Once each party has completed its Data Dictionary, URS/D&M and WA will exchange dictionaries. In addition, a sample data set to use in developing the Cross-reference Structure will be exchanged. The sample data set will include a representative cross section of all the data in the databases to ensure that significant data exchange problems are identified. The specific contents of the sample data set will be agreed upon by the Database Managers from the two parties. The Data Dictionaries and sample data set will be submitted by and to the respective Project Managers.

The contents of the Cross-reference Structures will not be defined until the Data Dictionaries are completed and the sample data set is provided to WA. However, it is likely that some differences exist between the two databases in terms of how the data are grouped and/or interpreted (e.g., differences in the way databases are normalized). Problems may also occur if definitions in the Database Dictionary are understood to mean different things to different users. The Cross-reference Structures will address these issues.

Using the Data Dictionaries and the sample data set, the WA and URS/D&M Database Managers will develop Draft Cross-reference Structures. The format and content of the Cross-reference Structures will be left up to the individual parties, since they will be developed solely for that party's use. The Draft Cross-reference Structures will be exchanged and reviewed by the other party to ensure that the database structures were interpreted correctly, and any needed changes will be incorporated into the Final Cross-reference Structures.

5. DATA SHARING

Once the final database structures have been established and tested, data exchange can begin. Data exchange requests may originate from either party. Occasions may arise when data transfers are needed from ongoing collection activities. In these instances, a routine data exchange schedule may be developed. The following sections describe the exchange process.

5.1 Points of Contact

The data request process will be initiated by a written request from/to the WA/URS/D&M Project Manager. In general, the primary point of contact at URS/D&M and WA for data transfer issues is the Database Manager at each company. The Database Managers will handle the technical aspects of data transfer with as-needed assistance from other individuals involved in the end use of transferred data. If a data transfer difference between URS/D&M and WA cannot be quickly resolved by the Database Managers and those providing assistance, then a data transfer meeting should be scheduled between the companies.

5.2 Data Requests

The data exchange process will be initiated by a written request from the WA or URS/D&M Project Manager in the form of an e-mail or hard copy specifying the samples, analytes, and fields pertinent to the request. The request should also state the due date for the Electronic Data Deliverable (EDD). These requests should be addressed to the Project Manager for the other party, who will evaluate the request and discuss it with their Database Manager. If the Project Manager for the delivering party has questions about the request or determines that delivering part or all of the requested data by the due date will not be possible, he/she will notify the receiving party by phone or send a written explanation as soon as possible. No data should be exchanged until it has been fully validated and approved. Special requests of a nature not documented here will be handled on a case by case basis.

5.3 Routine Data Exchange

The type of data designated for transfer may warrant a routine schedule for data exchange. In this case, data will be exchanged at an agreed-upon interval in the form of an EDD on CD or other electronic exchange medium. **A hard copy printout of the data must accompany the electronic files.** A schedule for these routine exchanges will be developed by the WA Database Manager and distributed to the WA Project Manager, the URS/D&M Project Manager, and the URS/D&M Database Manager. These EDDs will contain the relevant data that has been fully validated up

through that date. When the data from the new EDD are imported the data can be appended to the existing database.

5.4 Data Structure Changes

Both parties should make every effort to ensure that they conform to their established Data Dictionaries. Should unavoidable changes be made to the data structure of a database, exchanged data will still have to be sent in the original structure, or an addition to the Data Dictionary will need to be provided outlining the differences between the old and new structures. This insert to the Data Dictionary will be sent along with the respective data and a cover letter explaining the reasons for the structure change.

5.5 Exchange Medium

The data exchange medium is the electronic form that data are stored on while they are being shipped from one location to the other. Data exchange media may include diskettes, CDs, or alternative electronic media. The Database Manager for both companies must acknowledge they are capable of using an alternative electronic media before shipping an EDD using that media.

The data exchange media should be labeled with the date it was copied from the master database, the sending party name, and the receiving party name, and should be sent in a single package to the requestor. If the data cannot be stored entirely on one diskette, CD, or alternative medium, then the media should be numbered. **A hard copy printout of all included data must accompany the electronic exchange media.**

6. IMPLEMENTATION SCHEDULE

Implementation of the data integration process will begin upon approval of this Memorandum from both parties (DOE and UC Davis) and from the remedial project managers (RPMs). **By the end of July 2000, WA and URS/D&M will have discussed data needs, developed the Data Dictionaries, exchanged the sample data set, and created the Cross-reference Structures.** Following this initial period, data exchanges will occur on an as-needed basis. Integration of most or all of the data in the two databases will occur during preparation of the Site-wide Risk Assessment, scheduled to be completed in 2003.

7. REFERENCES

Dames & Moore, 1994, Final Draft RI/FS Work Plan, LEHR Environmental Restoration, University of California at Davis, 2 volumes.

U.S. Environmental Protection Agency, 1989, Risk Assessment Guidance for Superfund, Volume 1, Human Health Evaluation Manual (Part A), Interim Final, EPA540/1-89/002.

FIGURE

APPENDIX A

PROPOSED OUTLINE FOR DATA DICTIONARY

Proposed Outline for Data Dictionary

- 1 Introduction
 - 1.1 Software
 - 1.2 History of Database
 - 1.3 Size of Database

- 2 Relational Information
 - 2.1 Identification of tables
 - 2.2 Definition of table relationships
 - 2.3 Field definitions
 - 2.4 Field constraints and character

- 3 Data Characteristics
 - 3.1 Sample naming conventions
 - 3.2 Additional sample information
 - 3.2.1 Area of contamination
 - 3.2.2 Field location
 - 3.2.3 Sample medium
 - 3.2.4 Sample collection types and collection data
 - 3.2.5 Sample originator
 - 3.3 Analyte group naming conventions
 - 3.4 CAS names and numbers
 - 3.5 Analytical methods
 - 3.6 Units of measurement
 - 3.7 Detection limits
 - 3.8 Sample analysis information
 - 3.8.1 Sample analysis dates
 - 3.8.2 Dilution factors
 - 3.9 Data validation flags
 - 3.10 Duplicates, surrogates, matrix spikes, and blanks
 - 3.11 Level of data reliability

- 4 Data Manipulation / Data Flow
 - 4.1 Data flow diagram for data management
 - 4.2 Step by step data flow
 - 4.3 Miscellaneous data management activities
 - 4.4 Miscellaneous data validation activities